Sign in

Google Web Images Groups News Froogle Maps more »

rtf dtd Search Advanced Search Preferences

Web

Results 1 - 10 of about 748,000 for rtf dtd. (0.50 seconds)

Did you mean: *rdf* dtd

RTF FormattingKit FAQ
The RFK file extension is used for XML files that are valid according to the **RTF** FormattingKit **DTD**. If you open one of these files using an XML editor, ...
www.schema.de/eds/en/hierarchy-onClick.st4w;jsessionid=B9BF67415365779BF4E7FC5ACCB58805?c2VsZWN0ZWQ9NjY5N... - 29k - Cached - Similar pages

Balise 4.0 Overview
The **RTF** document scanner generates "ESIS" events corresponding to an implicit "**RTF DTD**", which models every logical piece of information one can find in an ...
xml.coverpages.org/balise40descr.html - 12k - Cached - Similar pages

Re: multiple **rtf** files
I think this was because > our **dtd** is very big; using xml might have helped ... end of file F:\FTEMP>jade -t **rtf** -cl:\jade\jadecurr\catalog -w no-valid -f ...
www.biglist.com/lists/dssslist/ archives/199802/msg00081.html - 7k - Cached - Similar pages

Re: XML to **RTF** transformation
Subject: Re: XML to **RTF** transformation ... you transformed your XML to tags that used an **RTF DTD**, you could use www.sema.be/mtc/products/rdc2rtf/index.html ...
www.oxygenxml.com/archives/ xsl-list/199905/msg00354.html - 9k - Cached - Similar pages

[DOC] NIMAS Conversion House Identification Form
File Format: Microsoft Word - View as HTML
We use an **RTF DTD** that works with JAWS. We are also capable of producing the VBK format which we believe is also DAISY and JAWS compliant. ...
nimas.cast.org/downloads/conversion_services/ NIMAS_ConvHouse_ICC_Macmillan.doc - Similar pages

Re: XML to **RTF** transformation
Re: XML to **RTF** transformation. ... transformed your XML to tags that used an **RTF DTD**, you could use www.sema.be/mtc/products/rdc2rtf/index.html there's also ...
www.stylusstudio.com/xsllist/199905/post40350.html - 15k - Cached - Similar pages

Re: bibiography **dtd**?
You want to map the names o in a source namespace (**RTF**) o to a target namespace (**DTD**). The source is the collection of instances, or documentation of ...
www.stylusstudio.com/xmldev/199911/post20260.html - 20k - Cached - Similar pages
[More results from www.stylusstudio.com]

[PPT] FreeBSD Documentation Project
File Format: Microsoft Powerpoint - View as HTML
RTF DTD -> **DTD** transformation; TeX. Formatted output: DSSSL. Stylesheets of choice: Norm Walsh's extensible DocBook stylesheets; Many knobs to turn and ...
people.freebsd.org/~nik/freebsdcon-1999/fdp/fdp.ppt - Similar pages

Logictran - R2Net FAQ
Then load in a sample **RTF** and use the "Translate **RTF**" toolbar icon. Two XML files are produced by default, conforming to the Docbook **DTD**. ...
www.logictran.net/support/r2net/faq.html - 39k - Cached - Similar pages

RATFINK
OSF's ``Rainmaker'' software can convert **RTF** documents to the Rainbow SGML document type; see ftp://ftp.ebt.com/pub/nv/**dtd**/rainbow/ ...
www.flightlab.com/cost/ratfink/ratfink.html - 54k - Cached - Similar pages

Did you mean to search for: ***rdf*** dtd

Try your search again on Google Book Search

Gooooooooogle ▶
Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**

rtf dtd Search

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve

Google Home - Advertising Programs - Business Solutions - About Google

©2006 Google

PORTAL
USPTO
Subscribe (Full Service) Register (Limited Service, Free) Login
Search: The ACM Digital Library The Guide
malformed <paragraph> document <paragraph> pdf
SEARCH
THE ACM DIGITAL LIBRARY
Feedback Report a problem Satisfaction survey

Terms used **malformed paragraph document paragraph pdf** Found **12,509** of **178,880**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window
Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next
Best 200 shown Relevance scale

1 Document formatting: Creating reusable well-structured PDF as a sequence of component object graphic (COG) elements
Steven R. Bagley, David F. Brailsford, Matthew R. B. Hardy
November 2003 **Proceedings of the 2003 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(458.01 KB) Additional Information: full citation, abstract, references, citings, index terms

Portable Document Format (PDF) is a page-oriented, graphically rich format based on PostScript semantics and it is also the format interpreted by the Adobe Acrobat viewers. Although each of the pages in a PDF document is an independent graphic object this property does not necessarily extend to the components (headings, diagrams, paragraphs etc.) within a page. This, in turn, makes the manipulation and extraction of graphic objects on a PDF page into a very difficult and uncertain process.The wo ...

Keywords: PDF, form Xobjects, graphic objects, tagged PDF

2 Document creation I: Creating structured PDF files using XML templates
Matthew R. B. Hardy, David F. Brailsford, Peter L. Thomas
October 2004 **Proceedings of the 2004 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(166.87 KB) Additional Information: full citation, abstract, references, index terms

This paper describes a tool for recombining the logical structure from an XML document with the typeset appearance of the corresponding PDF document. The tool uses the XML representation as a template for the insertion of the logical structure into the existing PDF document thereby creating a Structured/Tagged PDF. The addition of logical structure adds value to the PDF in three ways: the accessibility is improved (PDF screen readers for visually impaired users perform better) media options a ...

Keywords: PDF, XML, logical structure insertion

3 Information extraction and text segmentation: Structural extraction from visual layout of documents
Binyamin Rosenfeld, Ronen Feldman, Yonatan Aumann
November 2002 **Proceedings of the eleventh international conference on Information and knowledge management**
Publisher: ACM Press
Full text available: pdf(337.88 KB) Additional Information: full citation, abstract, references, index terms

Most information extraction systems focus on the textual content of the documents. They treat documents as sequences or of words, disregarding the physical and typographical layout of the information.. While this strategy helps in focusing the extraction process on the key semantic content of the document, much valuable information can also be derived form the document physical appearance. Often, fonts, physical positioning and other graphical characteristics are used to provide additional conte ...

4 Structure and transformation of documents: Mapping and displaying structural transformations between XML and PDF

Matthew R. B. Hardy, David F. Brailsford

November 2002 **Proceedings of the 2002 ACM symposium on Document engineering**

Publisher: ACM Press

Full text available: pdf(439.03 KB) Additional Information: full citation, abstract, references, citings, index terms

Documents are often marked up in XML-based tagsets to delineate major structural components such as headings, paragraphs, figure captions and so on, without much regard to their eventual displayed appearance. And yet these same abstract documents, after many transformations and 'typesetting' processes, often emerge in the popular format of Adobe PDF, either for dissemination or archiving.Until recently PDF has been a totally display-based document representation, relying on the underlying PostSc ...

Keywords: PDF, XML, document structure transformation

5 The LATEX legacy: 2.09 and all that

Chris Rowley

August 2001 **Proceedings of the twentieth annual ACM symposium on Principles of distributed computing**

Publisher: ACM Press

Full text available: pdf(956.36 KB) Additional Information: full citation, abstract, references, index terms

The second edition of The Manual [23] begins: `LATEX is a system for typesetting documents. Its first widely available version, mysteriously numbered 2.09, appeared in 1985.'

It is too early for a complete critical assessment of the impact of LATEX 2.09 because its world-wide effects on many aspects of many cultures, not least scientific publication, remain strong after 15 years—and that itself is significant in a technological world where a mere 15 months of fame can make *an* ...

6 Document authoring, markup and manipulation 2: Enhancing composite digital documents using XML-based standoff markup

Peter L. Thomas, David F. Brailsford

November 2005 **Proceedings of the 2005 ACM symposium on Document engineering DocEng '05**

Publisher: ACM Press

Full text available: pdf(695.86 KB) Additional Information: full citation, abstract, references, index terms

Document representations can rapidly become unwieldy if they try to encapsulate all possible document properties, ranging from abstract structure to detailed rendering and layout.We present a composite document approach wherein an XML-based document representation is linked via a 'shadow tree' of bi-directional pointers to a PDF representation of the same document. Using a two-window viewer any material selected in the PDF can be related back to the corresponding material in the XML, and vice ve ...

Keywords: MathML, MusicXML, PDF, XBL, XML, composite documents, standoff markup

7 Browsers and UI, web engineering, hypermedia & multimedia, security, and

accessibility: Using graph matching techniques to wrap data from PDF documents
Tamir Hassan, Robert Baumgartner
May 2006 **Proceedings of the 15th international conference on World Wide Web WWW '06**
Publisher: ACM Press
Full text available: pdf(191.28 KB) Additional Information: full citation, abstract, references, index terms

Wrapping is the process of navigating a data source, semi-automatically extracting data and transforming it into a form suitable for data processing applications. There are currently a number of established products on the market for wrapping data from web pages. One such approach is *Lixto* [1], a product of research performed at our institute.Our work is concerned with extending the wrapping functionality of Lixto to PDF documents. As the PDF format is relatively unstructured, this is a c ...

Keywords: PDF, document understanding, graph matching, logical structure, wrapping

8 Document authoring, markup and manipulation 1: Injecting information into atomic units of text
Yannis Haralambous, Gábor Bella
November 2005 **Proceedings of the 2005 ACM symposium on Document engineering DocEng '05**
Publisher: ACM Press
Full text available: pdf(244.01 KB) Additional Information: full citation, abstract, references, index terms

This paper presents a new approach to text processing, based on textemes. These are atomic text units generalising the concepts of character and glyph by merging them in a common data structure, together with an arbitrary number of user-defined properties. In the first part, we give a survey of the notions of character and glyph and their relation with Natural Language Processing models, some visual text representation issues and strategies adopted by file formats (SVG, PDF, DVI) and software (U ...

Keywords: OpenType, PDF, SVG, Unicode, character, glyph, multilingual typesetting, omega, texteme

9 ViSWeb—the Visual Semantic Web: unifying human and machine knowledge representations with Object-Process Methodology
Dov Dori
May 2004 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 13 Issue 2
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(1.22 MB) Additional Information: full citation, abstract, index terms

The Visual Semantic Web (ViSWeb) is a new paradigm for enhancing the current Semantic Web technology. Based on Object-Process Methodology (OPM), which enables modeling of systems in a single graphic and textual model, ViSWeb provides for representation of knowledge over the Web in a unified way that caters to human perceptions while also being machine processable. The advantages of the ViSWeb approach include equivalent graphic-text knowledge representation, visual navigability, semantic sentenc ...

Keywords: Conceptual graphs, Knowledge representation, Object-Process Methodology, Semantic Web, Visual Semantic Web

10 Books and reading: A document corpus browser for in-depth reading
Eric Bier, Lance Good, Kris Popat, Alan Newberger
June 2004 **Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(164.61 KB) Additional Information: full citation, abstract, references, citings, index

terms

Software tools, including Web browsers, e-books, electronic document formats, search engines, and digital libraries are changing the way people read, making it easier for them to find and view documents. However, while these tools provide significant help with short-term reading projects involving small numbers of documents, they provide less help with longer-term reading projects, in which a topic is to be understood in depth by reading many documents. For such projects, readers must find and m ...

Keywords: bookplex, computer-aided reading, digital library, document management, spatial memory, visualization, zoomable user interface

11 Paper session IR-1 (information retrieval): XML retrieval: Generalized contextualization method for XML information retrieval

Paavo Arvola, Marko Junkkari, Jaana Kekäläinen

October 2005 **Proceedings of the 14th ACM international conference on Information and knowledge management CIKM '05**

Publisher: ACM Press

Full text available: pdf(410.31 KB) Additional Information: full citation, abstract, references, index terms

A general re-weighting method, called contextualization, for more efficient element ranking in XML retrieval is introduced. Re-weighting is based on the idea of using the ancestors of an element as a context: if the element appears in a good context -- good interpreted as probability of relevance -- its weight is increased in relevance scoring; if the element appears in a bad context, its weight is decreased. The formal presentation of contextualization is given in a general XML representation a ...

Keywords: Dewey ordering, XML, contextualization, re-weighting, semi-structured data, structural indices, structured documents

12 P8: Transforming documentation from the XML doctypes used for the apache website to DITA

Donald M. Leslie

October 2001 **Proceedings of the 19th annual international conference on Computer documentation**

Publisher: ACM Press

Full text available: pdf(1.67 MB) Additional Information: full citation, abstract, references, citings, index terms

A primary factor behind the enormous interest in XML is the support it provides for transforming documents to meet the needs of information-processing applications as well as human readers working with HTML, print, and other presentation media. This case study reviews the issues we confronted, the tools we implemented, and the procedures we adopted to transform a documentation set from one XML document type to another, and from XML to HTML and Adobe PDF.The documentation set for Xalan, the Apach ...

Keywords: PDF, XML, XSL, XSLT, document transformation, formatting objects, stylesheets

13 Structural computing: Unifying structure, behavior, and data with themis types and templates

William Van Lepthien, Kenneth M. Anderson

August 2004 **Proceedings of the fifteenth ACM conference on Hypertext and hypermedia HYPERTEXT '04**

Publisher: ACM Press

Full text available: pdf(332.32 KB) Additional Information: full citation, abstract, references, index terms

Structural computing evolved from work on open hypermedia to aid in the creation of

software infrastructure. Open hypermedia had produced software that provided applications with access to hypermedia structures and services. The question was asked if these results could be generalized to create similar tools for other domains. Initial work focused on the development of structure servers that can create and manipulate domain-specific structures, but little work focused on allowing those structure ...

Keywords: Chimera, Themis, structural computing, templates, types

14 Information retrieval models: Detecting similar documents using salient terms

James W. Cooper, Anni R. Coden, Eric W. Brown
November 2002 **Proceedings of the eleventh international conference on Information and knowledge management**
Publisher: ACM Press

Full text available: pdf(180.68 KB) Additional Information: full citation, abstract, references, citings, index terms

We describe a system for rapidly determining document similarity among a set of documents obtained from an information retrieval (IR) system. We obtain a ranked list of the most important terms in each document using a rapid phrase recognizer system. We store these in a database and compute document similarity using a simple database query. If the number of terms found to not be contained in both documents is less than some predetermined threshold compared to the total number of terms in the doc ...

Keywords: databases, document similarity, duplicate documents, shingles, text mining

15 Document creation II: Page composition using PPML as a link-editing script

Steven R. Bagley, David F. Brailsford
October 2004 **Proceedings of the 2004 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(197.33 KB) Additional Information: full citation, abstract, references, index terms

The advantages of a COG (Component Object Graphic) approach to the composition of PDF pages have been set out in a previous paper [1]. However if pages are to be composed in this way then the individual graphic objects must have known bounding boxes and must be correctly placed on the page in a process that resembles the link editing of a multi-module computer program. Ideally the linker should be able to utilize all declared resource information attached to each COG.

We have investiga ...

Keywords: PDF, PPML, form Xobjects, graphic objects, link editing

16 Innovative Document Systems: The multivalent browser: a platform for new ideas

Thomas A. Phelps, Robert Wilensky
November 2001 **Proceedings of the 2001 ACM Symposium on Document engineering**
Publisher: ACM Press

Full text available: pdf(188.51 KB) Additional Information: full citation, abstract, references, citings, index terms

The Multivalent Browser is built on a architecture that separates functionality from concrete document format. Almost all functionality is made available via relatively small modules of code called behaviors that programmers can write to extend the core system. Behaviors can be as significant and powerful as parser-renderers for scanned paper, HTML, or TeX DVI; as fine-grained as hyperlinks, cookies, and the disabling of menu items; and as innovative or uncommon as in situ annotatins, "lenses", ...

Keywords: annotation, architecture, digital, document, multivalent behavior, paper, scanned

17 Paper to HTML—an automatic, seamless process for documentation production

Virginie Ahrens, Valérie Lecompte
October 1999 **Proceedings of the 17th annual international conference on Computer documentation**
Publisher: ACM Press
Full text available: pdf(672.82 KB) Additional Information: full citation, abstract, index terms

This paper describes how ILOG, a French software company designing C++ and Java class libraries, managed the transition between paper-only documentation and extensive HTML online documentation in less than two years. In this paper, we analyze the underlying reasons for making this change, describe the technological choices that were made, and walk through the various steps of the project from its beginning to final completion.

Keywords: C++ and Java class libraries, HTML, Java Script, Web design, modularity, online documentation, page-authoring tools, portability, reusability

18 Document presentation: Support for arbitrary regions in XSL-FO

Ana Cristina B. da Silva, Joao B. S. de Oliveira, Fernando T. M. Mano, Thiago B. Silva, Leonardo L. Meirelles, Felipe R. Meneguzzi, Fabio Giannetti
November 2005 **Proceedings of the 2005 ACM symposium on Document engineering DocEng '05**
Publisher: ACM Press
Full text available: pdf(520.86 KB) Additional Information: full citation, abstract, references, index terms

This paper proposes an extension of the XSL-FO standard which allows the specification of an unlimited number of arbitrarily shaped page regions. These extensions are built on top of XSL-FO 1.1 to enable flow content to be laid out into arbitrary shapes and allowing for page layouts currently available only to desktop publishing software. Such a proposal is expected to leverage XSL-FO towards usage as an enabling technology in the generation of content intended for personalized printing.

Keywords: LaTeX, SVG, XML, XSL-FO, arbitrary shapes, digital publishing, typesetting

19 Add one egg, a cup of milk, and stir: single source documentation for today

Carl Stieren
October 1997 **Proceedings of the 15th annual international conference on Computer documentation**
Publisher: ACM Press
Full text available: pdf(776.88 KB) Additional Information: full citation, references, citings, index terms

20 Making use of document standards and models: A framework for structure, layout & function in documents

John Lumley, Roger Gimson, Owen Rees
November 2005 **Proceedings of the 2005 ACM symposium on Document engineering DocEng '05**
Publisher: ACM Press
Full text available: pdf(1.55 MB) Additional Information: full citation, abstract, references, index terms

The Document Description Framework (DDF) is a representation for variable-data documents. It supports very high flexibility in the type and extent of variation supported, considerably beyond the 'copy-hole' or flow-based mechanisms of existing formats and tools. DDF is based on holding application data, logical data struc-ture and presentation as well as constructional 'programs' together within a single document. DDF documents can be merged with other documents, bound to variable values increme ...

Keywords: SVG, XML, XSLT, document construction, functional programming

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

PORTAL
USPTO
Subscribe (Full Service) Register (Limited Service, Free) Login
Search: The ACM Digital Library The Guide
malformed <paragraph> document <paragraph> pdf
SEARCH
THE ACM DIGITAL LIBRARY
Feedback Report a problem Satisfaction survey

Terms used **malformed paragraph document paragraph pdf** Found **12,509** of **178,880**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 21 - 40 of 200 Result page: previous 1 **2** 3 4 5 6 7 8 9 10 next
Best 200 shown Relevance scale

21 Webbed documents

Malcolm Graham, Andrew Surray
February 1996 **Proceedings of the 13th annual international conference on Systems documentation: emerging from chaos: solutions for the growing complexity of our jobs**
Publisher: ACM Press
Full text available: pdf(513.06 KB) Additional Information: full citation, references, index terms

22 Tools & techniques track: applying machine learning to collection development: Developing practical automatic metadata assignment and evaluation tools for internet resources

Gordon W. Paynter
June 2005 **Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(305.60 KB) Additional Information: full citation, abstract, references, index terms

This paper describes the development of practical automatic metadata assignment tools to support automatic record creation for virtual libraries, metadata repositories and digital libraries, with particular reference to library-standard metadata. The development process is incremental in nature, and depends upon an automatic metadata evaluation tool to objectively measure its progress. The evaluation tool is based on and informed by the metadata created and maintained by librarian experts at the ...

Keywords: INFOMINE, automatic metadata assignment, automatic metadata evaluation, iVia, metadata

23 Combining optimal clustering and Hidden Markov models for extractive summarization

Pascale Fung, Grace Ngai, Chi-Shun Cheung
July 2003 **Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering - Volume 12**
Publisher: Association for Computational Linguistics
Full text available: pdf(140.12 KB) Additional Information: full citation, abstract, references

We propose Hidden Markov models with unsupervised training for extractive summarization. Extractive summarization selects salient sentences from documents to be included in a summary. Unsupervised clustering combined with heuristics is a popular approach because no annotated data is required. However, conventional clustering methods such as K-means do not take text cohesion into consideration. Probabilistic

methods are more rigorous and robust, but they usually require supervised training with a ...

24 Hypermedia and Graphics 2: Vector graphics: from PostScript and Flash to SVG

Steve Probets, Julius Mong, David Evans, David Brailsford

November 2001 **Proceedings of the 2001 ACM Symposium on Document engineering**

Publisher: ACM Press

Full text available: pdf(127.00 KB) Additional Information: full citation, abstract, references, citings, index terms

The XML-based specification for Scalable Vector Graphics(SVG), sponsored by the World Wide Web consortium, allows for compact and descriptive vector graphics for the Web.This paper describes a set of three tools for creating SVG, either from first principles or via the conversion of existing formats. The *ab initio* generation of SVG is effected from a server-side CGI script, using a PERL library of drawing functions; later sections highlight the problems of converting Adobe PostScript and ...

Keywords: Flash, PDF, PostScript, SVG, SWF

25 Industry track: Intelligent agent for automated manufacturing rule generation

Alan Clark, Dimitar Filev

November 2004 **Proceedings of the thirteenth ACM international conference on Information and knowledge management CIKM '04**

Publisher: ACM Press

Full text available: pdf(247.85 KB) Additional Information: full citation, references, index terms

Keywords: KBE, clustering algorithms, intelligent agent, knowledge extraction, knowledge management, latent semantic indexing

26 QARAB: a question answering system to support the Arabic language

Bassam Hammo, Hani Abu-Salem, Steven Lytinen

July 2002 **Proceedings of the ACL-02 workshop on Computational approaches to semitic languages**

Publisher: Association for Computational Linguistics

Full text available: pdf(339.08 KB) Additional Information: full citation, abstract, references

We describe the design and implementation of a question answering (QA) system called QARAB. It is a system that takes natural language questions expressed in the Arabic language and attempts to provide short answers. The system's primary source of knowledge is a collection of Arabic newspaper text extracted from Al-Raya, a newspaper published in Qatar. During the last few years the information retrieval community has attacked this problem for English using standard IR techniques with only medioc ...

27 SST: using single-sourcing, SGML, and teamwork for documentation

Carl Stieren

October 1999 **Proceedings of the 17th annual international conference on Computer documentation**

Publisher: ACM Press

Full text available: pdf(784.56 KB) Additional Information: full citation, abstract, references, index terms

Suppose you don't have a fancy database-driven system to generate your documentation. How can you develop single-source documentation for output in multiple formats, without having to store your source in a specific format that will soon become obsolete? The answer is to use a combination of your own SGML or XML tags to mark up your documentation and a simple OmniMark® program to create each output format and presentation style. There's also a third ingredient: teamwork. As much as any ...

Keywords: HTML, SGML, XML, print, single-source, teamwork

28 Document analysis: Visual signature based identification of Low-resolution document images
Ardhendu Behera, Denis Lalanne, Rolf Ingold
October 2004 **Proceedings of the 2004 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(2.00 MB) Additional Information: full citation, abstract, references, index terms

In this paper, we present (a) a method for identifying documents captured from low-resolution devices such as web-cams, digital cameras or mobile phones and (b) a technique for extracting their textual content without performing OCR. The first method associates a hierarchically structured visual signature to the low-resolution document image and further matches it with the visual signatures of the original high-resolution document images, stored in PDF form in a repository. The matching algor ...

Keywords: document visual signature, document-based meeting retrieval, documents' content extraction, low-resolution document image identification

29 Technical session 10: watermarking and multi-media processing: Thematic segmentation of meetings through document/speech alignment
Dalila Mekhaldi, Denis Lalanne, Rolf Ingold
October 2004 **Proceedings of the 12th annual ACM international conference on Multimedia**
Publisher: ACM Press
Full text available: pdf(917.84 KB) Additional Information: full citation, abstract, references, citings, index terms

This article proposes a multimodal approach for segmenting meeting recordings. This bi-modal method takes advantages of the alignment of speech transcript with documents, in the context of meetings or lectures, where documents are discussed. The method first displays the alignment results as a set of nodes in a 2D space, where the two axes represent respectively the documents content and the speech transcript. The most connected regions in this graph are detected using a clustering method. Th ...

Keywords: clustering techniques, document analysis, meeting dialogs structuring, multimedia information retrieval, multimodal thematic alignment, thematic segmentation

30 Document access and understanding: Methods for the semantic analysis of document markup
Petra Saskia Bayerl, Harald Lüngen, Daniela Goecke, Andreas Witt, Daniel Naber
November 2003 **Proceedings of the 2003 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(230.61 KB) Additional Information: full citation, abstract, references, index terms

We present an approach on how to investigate what kind of semantic information is regularly associated with the structural markup of scientific articles. This approach addresses the need for an explicit formal description of the semantics of text-oriented XML-documents. The domain of our investigation is a corpus of scientific articles from psychology and linguistics from both English and German online available journals.For our analyses, we provide XML-markup representing two kinds of semantic ...

Keywords: XML, information extraction, prolog, semantic analysis

31 Document management: A ground-truthing engine for proofsetting, publishing, re-purposing and quality assurance

Steven J. Simske, Margaret Sturgill
November 2003 **Proceedings of the 2003 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(165.66 KB) Additional Information: full citation, abstract, references, index terms

We present design strategies, implementation preferences and throughput results obtained in deploying a UI-based ground truthing engine as the last step in the quality assurance (QA) for the conversion of a large out-of-print book collection into digital form. A series of automated QA steps were first performed on the document. Five distinct zoning analysis options were deployed and the PDF output thence generated was used to regenerate TIFF files for comparison to the originals. Regenerated TIF ...

Keywords: layout, print-on-demand, region management, templates

32 Software engineering and the Internet: a roadmap
Luca Bompani, Paolo Ciancarini, Fabio Vitali
May 2000 **Proceedings of the Conference on The Future of Software Engineering**
Publisher: ACM Press
Full text available: pdf(1.73 MB) Additional Information: full citation, references, citings, index terms

33 Research session: new applications: The SphereSearch engine for unified ranked retrieval of heterogeneous XML and web documents
Jens Graupmann, Ralf Schenkel, Gerhard Weikum
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment
Full text available: pdf(381.86 KB) Additional Information: full citation, abstract, references, index terms

This paper presents the novel SphereSearch Engine that provides unified ranked retrieval on heterogeneous XML and Web data. Its search capabilities include vague structure conditions, text content conditions, and relevance ranking based on IR statistics and statistically quantified ontological relationships. Web pages in HTML or PDF are automatically converted into XML format, with the option of generating semantic tags by means of linguistic annotation tools. For Web data the XML-oriented query ...

34 Extracting mathematical expressions from postscript documents
Michael Yang, Richard Fateman
July 2004 **Proceedings of the 2004 international symposium on Symbolic and algebraic computation**
Publisher: ACM Press
Full text available: pdf(137.43 KB) Additional Information: full citation, abstract, references, index terms

Full-text indexing of documents containing mathematics cannot be considered a complete success unless the mathematics symbolism is extracted and represented in a standardized form permitting both searching for formulas, and re-use of this information in (for example) computer algebra systems. Most documents produced in the past and subsequently digitally encoded, and even most of those potentially "born digital" in current journal production are---at best---encoded in a printer form such as Adob ...

Keywords: digital library, document image analysis, mathematics, optical character recognition, postscript documents

35 E-rulemaking: Near-duplicate detection for eRulemaking
Hui Yang, Jamie Callan
May 2005 **Proceedings of the 2005 national conference on Digital government research dg.o2005**

Publisher: Digital Government Research Center

Full text available: pdf(248.00 KB) Additional Information: full citation, abstract, references, index terms

U.S. regulatory agencies are required to solicit, consider, and respond to public comments before issuing regulations. In recent years, agencies have begun to accept comments via both email and Web forms. The transition from paper to electronic comments makes it much easier for individuals to customize "form" letters, which they do, creating "near-duplicate" comments that express the same viewpoint in slightly different languages. This paper explores the use of simple text clustering and retriev ...

Keywords: eRulemaking, information retrieval, near duplicate detection, public comments

36 A survey and analysis of Electronic Healthcare Record standards

Marco Eichelberg, Thomas Aden, Jörg Riesmeier, Asuman Dogac, Gokce B. Laleci

December 2005 **ACM Computing Surveys (CSUR)**, Volume 37 Issue 4

Publisher: ACM Press

Full text available: pdf(844.11 KB) Additional Information: full citation, abstract, references, index terms

Medical information systems today store clinical information about patients in all kinds of proprietary formats. To address the resulting interoperability problems, several Electronic Healthcare Record standards that structure the clinical content for the purpose of exchange are currently under development. In this article, we present a survey of the most relevant Electronic Healthcare Record standards, examine the level of interoperability they provide, and assess their functionality in terms o ...

Keywords: Electronic Healthcare Record standards, eHealth, interoperability

37 Unifying documentation teams

Karl A. Hakkarainen

October 1999 **Proceedings of the 17th annual international conference on Computer documentation**

Publisher: ACM Press

Full text available: pdf(1.11 MB) Additional Information: full citation, abstract, references, citings, index terms

This paper describes how a set of geographically and organizationally distributed documentation teams created the Rational Suite 1.0 documentation set. The paper covers the business operations of Rational Software, details the documentation tools and technologies used in the project and describes the evolution of the larger team as it learned how to work with a new software development methodology. The paper concludes with a summary of lessons learned and next steps.

Keywords: distributed teams, documentation standards, organizational development

38 Document interaction: ScreenCrayons: annotating anything

Dan R. Olsen, Trent Taufer, Jerry Alan Fails

October 2004 **Proceedings of the 17th annual ACM symposium on User interface software and technology**

Publisher: ACM Press

Full text available: pdf(586.83 KB) Additional Information: full citation, abstract, references, index terms

ScreenCrayons is a system for collecting annotations on any type of document or visual information from any application. The basis for the system is a screen capture upon which the user can highlight the relevant portions of the image. The user can define any number of topics for organizing notes. Each topic is associated with a highlighting "crayon." In addition the user can supply annotations in digital ink or text. Algorithms are described that summarize captured images based on the highli ...

Keywords: annotation, digital ink, image summarization, screen capture

39 Papers: collaborating through documents: Moving markup: repositioning freeform annotations

Gene Golovchinsky, Laurent Denoue
October 2002 **Proceedings of the 15th annual ACM symposium on User interface software and technology**
Publisher: ACM Press

Full text available: pdf(576.12 KB) Additional Information: full citation, abstract, references, citings, index terms

Freeform digital ink annotation allows readers to interact with documents in an intuitive and familiar manner. Such marks are easy to manage on static documents, and provide a familiar annotation experience. In this paper, we describe an implementation of a freeform annotation system that accommodates dynamic document layout. The algorithm preserves the correct position of annotations when documents are viewed with different fonts or font sizes, with different aspect ratios, or on different devi ...

Keywords: annotation, dynamic document layout, freeform digital ink, repositioning annotations

40 Demonstrations: A demonstration of the document description framework

John Lumley, Roger Gimson, Owen Rees
November 2005 **Proceedings of the 2005 ACM symposium on Document engineering DocEng '05**
Publisher: ACM Press

Full text available: pdf(20.86 KB) Additional Information: full citation, abstract, references, index terms

The Document Description Framework (DDF) [1] is a representation for variable-data documents, designed to support very high flexibility in the type and extent of variation, considerably beyond 'copy-hole' or flow-based mechanisms of existing formats and tools. This demonstration shows how i) DDF documents can be evaluated and merged to construct complex multi-stage documents, ii) the layout capabilities can be extended flexibly and iii) how they may be created and edited within a GUI-based envir ...

Keywords: SVG, XML, XSLT, document construction, functional programming

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.
Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat · QuickTime Windows Media Player Real Player